DOCUMENT RESUME

ED 268 181 TM 860 248

AUTHOR Fuchs, Lynn S.
TITLE Review of Monitoring Procedures with Mildly Handicapped Students.
PUB DATE [86]
NOTE 36p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Competency Based Education; Educational Assessment; Educational Improvement; Educational Methods; *Evaluation Methods; *Formative Evaluation; Measurement Objectives; Mild Disabilities; Remedial Instruction; *Research Methodology; *Special Education; *Success
IDENTIFIERS *Monitoring

ABSTRACT

This paper presents a comprehensive review of special and remedial education practice in formative evaluation of goal attainment among mildly handicapped students. First, the relationship of formative evaluation to the range of special education assessment practice and its importance to the field are described. Then, four critical issues in formative evaluation methodology are discussed: focus of measurement, frequency of measurement, data display, and data-utilization methods. The reviewed research supports the use of ongoing criterion-referenced monitoring systems to improve the instructional programs of mildly handicapped students. Research also provides the basis for several statements on the nature of effective monitoring systems among mildly handicapped populations. Specifically, the data base supports the use of (1) certain empirically validated critical behaviors as the focus of monitoring activity; (2) long-term goal statements that may encourage teachers to focus not only on the immediate instructional content, but also on maintenance and generalization of skills; and (3) relatively ambitious goals that support task persistence and striving. The literature also provides a rationale for teachers to measure student performance at least twice weekly; to graph data using their preferred convention; and to employ data-utilization rules for determining when, and perhaps how, to modify students' instructional programs. An eight-page reference list is appended. (PN)

****
* Reproductions supplied by EDRS are the best that can be made . *
* from the original document. *
****

ED268181

Review of Monitoring Procedures with Mildly Handicapped Students

Lynn S. Fuchs

George Peabody College

Vanderbilt University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. S. Fuchs

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Requests for reprints should be sent to: Lynn S. Fuchs, Department of Special Education, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203.

Running Head: Monitoring Procedures

TM 860 248

Abstract

The purpose of this paper was to review special education practice in formative evaluation of goal attainment among mildly handicapped pupils. First, the relationship of formative evaluation to the range of special education assessment practice and its importance to the field are described. Then, four critical issues in formative evaluation methodology are discussed: focus of measurement, frequency of measurement, data display, and data-utilization methods. Finally, a proposal is advanced for additional related research.

Review of Monitoring Procedures with Mildly Handicapped Students

The specification of goals and the evaluation of goal attainment are fundamental to American schooling. Historically, primary goals of public education have been those of Americanization and instilling democratic values in youth (Mulhern, 1959), and evaluation has been conducted in a routinely clerical way through accumulation of enrollment and attendance records (Campbell, Cunningham, Nystrand, & Usdan, 1975). Over time, however, notions concerning the nature of useful educational goals and what activities constitute effective evaluation have changed dramatically. The nature of this evolution is reviewed briefly below.

In the nineteenth century, educational goals and evaluation were broad and related marginally to academic curriculum. Psychologists viewed the brain as a composite of general intellectual activity (Eisner, 1967), and, relatedly, prevalent beliefs held that identifying and strenghthening general faculties would produce concurrent educational growth. This conceptualization of learning fostered the global definition and evaluation of educational outcomes.

At the turn of the century, however, notable movement toward specificity in goal definition and evaluation occurred as psychologists began to develop the notion that educational growth might be operationalized in terms of series of specific learning products. This notion was based on Thorndike's work, which demonstrated the specificity of transfer wherein generalization of learning occurs when elements in the original learning context are relevant and similar to elements in other contexts (Eisner, 1967). Applying Thorndike's work to the development of educational curricula, Bobbitt (cited in Eisner, 1967) argued that life consists of the

performance of specific activities and that the numerous and discrete skills and knowledge required for successful adulthood should constitute the curriculum of schools.

This premise was central to Ralph Tyler's work in curriculum and evaluation. As director of the National Assessment Project, he required that specific educational objectives constitute the basis for designing evaluation instruments. Under the auspices of the National Assessment of Educational Progress, there was sustained effort to specify distinct academic goals and to develop an array of related psychometrically adequate, group achievement tests. Standardized administration of these tests to large groups of students created a national data base by which summative comparisons of educational effectiveness could be formulated among regions of the country, school systems, schools, and pupils of varying characteristics. This development represented a critical move toward more direct, standardized measurement of group academic achievement as a summative index of program effectiveness and goal attainment.

Psychologists like Gagne, Glaser, and Mager also were interested in developing clear statements of educational objectives and achievement tests based on those objectives (Bloom, Madaus, & Hastings, 1971). Compared to Tyler, however, these researchers focused more on development of effective instruction than on summative evaluation. They proposed that, in order to increase educational effectiveness, educators should measure pupil achievement over time, in relation to a specific set of desired outcome behaviors. This idea led to the development of a methodology of criterion-referenced formative evaluation.

Criterion-referenced formative evaluation is the ongoing collection of pupil performance data, during program implementation, and with respect to mastery of behaviorally-stated goals. The purpose of such data

collection is to generate an information base that informs decisions concerning how to revise and improve programs. Consequently, formative evaluation methodology addresses strategies for (a) specifying clear, distinct instructional objectives or domains (e.g., Bloom et al., 1971; Popham, 1980), and (b) developing technically adequate criterion-referenced measurement procedures that can be matched isomorphically to instructional objectives or domains (e.g., Popham, 1980).

Formative evaluation for revision and improvement of instructional programs perhaps has realized its greatest impact in special and remedial education programs, where conventional instructional practice, by definition, is ineffective. The literature on the effectiveness of formative evaluation, or continuous monitoring of individual pupil progress with revision of educational programs, is robust: The Direct Instruction (e.g., Gersten, Carnine, & White, 1984), applied behavior analysis (e.g., Lovitt, 1981; Rieth, Polsgrove, & Semmel, 1981), general special education (e.g., Fuchs, Deno, & Mirkin, 1984), effective schools (Eubanks & Levine, 1983; Hoffman & Rutherford, 1984) and special eduation effective teaching (e.g., Goodman, 1985; Peterson, Albert, Foxworth, Cox, & Tilley, 1985) literatures all include ongoing objective-referenced monitoring of pupil progress as an essential component of effective teaching.

Nevertheless, a comprehensive review of special and remedial education practice in formative evaluation of goals is lacking; and without such integration of available empirical work, it remains unclear how practitioners can implement formative evaluation most effectively. Therefore, the purpose of this paper was to conduct such a review. This integration is organized into three major sections. In the first part, the relationship of formative evaluation to the spectrum of special education assessment practice and its importance to the field are described. Then,

four critical issues in formative evaluation methodology are discussed, with a review of recent research in each area. These critical issues concern the form and frequency of assessment and methods of data display and data utilization. Finally, a proposal is presented for future research.

Importance of Monitoring as a Special Education Assessment Activity

According to Salvia and Ysseldyke (1985), special education assessment is the process of collecting data in order to specify and verify problems and to formulate decisions about pupils with respect to referral, screening, classification, instructional planning, and program modification. Decisions in the first three assessment phases constitute the identification process, wherein norm-referenced comparisons among pupils are made to judge whether students are sufficiently discrepant from peers to require special intervention.

Decisions in the remaining assessment phases, instructional planning and program modification, are related more integrally to instructional program content and methodology. In the first of these two phases, specific problems are identified for educational intervention, student characteristics are described, the educational ecology is assessed, and an initial hypothesis concerning instruction is generated. In the second phase, the effectiveness of the instructional hypothesis is evaluated through ongoing measurement of pupil progress, and the cycle of postulating and testing instructional hypotheses continues.

Theoretically, the instructional planning and program modification phases of assessment complement one another. Nevertheless, in practice, they have become associated with markedly different and, for the most part, mutually exclusive approaches to the development of special education instructional programs. The first approach, aptitude-treatment interaction (ATI), embodies the instructional planning phase: the initial description of

learners, wherein student aptitudes are presumed to interact predictably with instructional treatments to produce comparatively strong student learning. Thus, with an ATI approach, the development or selection of educational programs is deductive, derived from prior explication of learner characteristics.

The second approach is formative evaluation or ongoing objective-referenced monitoring of pupil progress. Whereas an ATI approach emphasizes the importance of the first phase of program planning, describing salient learner characteristics, formative evaluation embodies program planning's second major phase, ongoing evaluation and modification of proposed programs, wherein student performance is measured repeatedly under different instructional conditions. The purpose of this measurement is to provide a data base with which effective instructional programs may be developed empirically. Thus, formative evaluation is an inductive, rather than deductive, approach to developing instructional plans.

An ATI approach represents the traditional and prevalent method for formulating educational programs (e.g., Ysseldyke & Thurlow, 1984). Nevertheless, important problems have been associated with ATI-related methodology, and formative evaluation appears more tenable for several related reasons. First, formative evaluation avoids ATI's reliance on initial diagnoses of learner characteristics for prescribing treatment when (a) conceptualizations of cognitive abilities are incomplete (Ysseldyke, 1979), (b) available tests of pupil characteristics are psychometrically inadequate (Arter & Jenkins, 1979; Coles, 1978; Glaser, 1981; Salvia & Ysseldyke, 1985), and (c) the nature of interactions among learner and teacher characteristics, educational treatments, and classroom environments, to a large extent, is unknown (Ysseldyke, 1979). Second, the formative evaluation related practice of repeated measurement by classroom teachers in

classroom settings is more ecologically valid and less reactive than the use of traditional assessment procedures associated with typical ATI approaches (D. Fuchs & L. Fuchs, in press a). Furthermore, it avoids traditional testing procedures associated with ATI methodology (e.g., one test session administered by an unfamiliar examiner), which may discriminate systematically against handicapped, minority, and low socioeconomic students (D. Fuchs & L. Fuchs, in press b; Fuchs, Fuchs, Power, & Dailey, 1985).

Perhaps most importantly, however, research on the effectiveness of formative evaluation is more promising than the literature associated with ATI approaches. In a recent meta-analysis of the effects on student achievement of formative evaluation (L. Fuchs & D. Fuchs, in press), the average effect size was .70. This indicates one can expect students whose programs are monitored systematically and developed formatively over time to achieve, on average, seven-tenths of a standard deviation unit higher than students whose programs are not monitored systematically and developed formatively. In terms of the standard normal curve and an achievement test scale with a population mean of 100 and standard deviation of 15, the use of formative evaluation to generate and evaluate individualized programs can be expected to raise the typical achievement outcome score from 100.00 to 110.50, or from the 50th to the 76th percentile.

The use of formative evaluation, then, appears to increase academic achievement reliably, and this program planning methodology is available for practitioners to inductively formulate instructionally-related assessment decisions and successful individual educational programs. This conclusion contrasts sharply with (a) the research base indicating that ATI approaches, specifically, fail to improve the achievement of handicapped learners (e.g., Arter & Jenkins, 1979; Hammill & Larsen, 1974; Hammill & Wiederholt, 1973), and (b) a body of literature that questions the effectiveness of special

education in general (e.g., Dunn, 1968; Glass, 1983).

Consequently, this alternative methodology for developing students' instructional programs may represent a critical component of special education assessment practice. Furthermore, its potential importance is highlighted by and implicitly recognized in The Education for All Handicapped Children Act of 1975, which mandates that individualized educational programs (IEPs) include evaluation procedures for assessing whether goals and objectives are being achieved. Criterion-referenced formative evaluation appears consonant with this Federal mandate and with public demand, in general, for accountability in the schools.

Although substantive compliance with this Federal legislation suggests that special educators routinely engage in criterion-referenced measurement for evaluation of goal attainment (Deno & Mirkin, 1977; Fuchs & Fuchs, 1984), research (Fuchs, Fuchs, & Warren, 1982) on the procedures special educators use to evaluate student mastery of IEP goals indicates that they formulate criterion-referenced decisions primarily on the basis of unsystematic observation, rather than on the basis of assessment data (Fuchs & Fuchs, 1984). Unfortunately, although teachers express confidence in the accuracy of those criterion-referenced judgments (Fuchs, Fuchs, & Warren, 1982), their informal evaluations about objective mastery tend to be inaccurate and to overrate student performance (Fuchs & Fuchs, 1984). This suggests the need for practitioners to design and implement criterion-referenced measurement systems for formulating valid criterion-referenced decisions concerning student progress. Below critical issues for teachers to consider in designing these systems are discussed.

Designing Criterion-Referenced Formative Evaluation Systems: Critical Dimensions

Analysis of the literature on criterion-referenced formative evaluation for mildly handicapped pupils indicates that certain elements of ongoing monitoring systems may be critical in effecting desired student achievement outcomes. As practitioners design procedures for measuring student progress toward goal attainment, four important elements are the focus of measurement, frequency of measurement, data display, and data-utilization methods. A discussion of research related to each dimension follows.

Focus of Measurement

In designing measurement systems, a first consideration is the focus of measurement, which involves specifying measurement and instructional goals. Three relevant dimensions of goal statements are: What simple, observable behaviors are critical indicators of student performance?; What is an appropriate breadth of goal statement?; and What principles are useful for determining mastery criteria?

Critical behaviors. A relatively new, but growing body of research concerns what critical behaviors can be used to monitor mildly handicapped students' progress in basic skills reliably, validly, and practically. This research provides practitioners with information concerning what behaviors are critical indicators of student growth and, relatedly, what behaviors might be useful to incorporate into basic skills goal statements.

Findings indicate that repeated 1 to 3 minute measurements of simple behaviors, such as reading aloud isolated words or passages (Deno, Mirkin, & Chiang, 1982), writing words in response to story starters (Deno, Marston, & Mirkin, 1982), and spelling words or letter sequences (Deno, Mirkin, Lowry, & Kuehnle, 1980), provide meaningful time series of academic performance. The measures demonstrate stability (Fuchs, Deno, & Marston, 1983), interscorer and alternate form reliabilty (Marston, 1982; Marston & Deno,

1981), sensitivity to student growth (Marston, Fuchs, & Deno, 1985), criterion validity with respect to contruct validated, widely accepted tests (Deno, Marston, & Mirkin, 1982; Deno, Mirkin, & Chiang, 1982; Deno et al., 1980) and teacher ratings (Marston et al., 1985; Fuchs, 1981), discriminative validity with respect to students' special education label (Deno, Marston, & Mirkin, 1982; Deno, Mirkin, & Chiang, 1982; Deno et al., 1980), and logistical feasibility for practitioners (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1981). These findings indicate which behaviors might be incorporated into goal statements and can be used to track student progress over time by applied behavior analysts, precision teachers, and other designers of monitoring systems.

Breadth of goal statements. In comparison to general concensus in the literature concerning what behaviors are useful for monitoring, greater controversy surrounds the issue of breadth of goal statement toward which progress should be assessed. Currently, practitioners can monitor progress toward one of two types of goal statements, each representing a different breadth: One focuses on attainment of long-term goals, the other on mastery of short-term objectives.

With a long-term goal approach, an annual goal is specified and a large pool of related measurement items is created. From this measurement pool, subsets of items, or monitoring probes, are drawn randomly (see Fuchs, Deno, & Mirkin, 1984), and the difficulty level of the monitoring probes remains constant over the year. Contrastingly, with a short-term objective approach, a series of objectives corresponding to steps within a hierarchical curriculum is specified, and a series of relatively circumscribed, small pools of items are created, each of which corresponds to a specific objective (see Lindsley, 1971; White & Haring, 1980). The difficulty level of material on which students are measured increases as

students master the sequentially-related objectives.

By definition, both types of measurement are ongoing, criterion-referenced, curriculum-based, and enjoy strong curricular validity or correspondence between tests and programmatic goals and objectives (McClung cited in Yalow & Popham, 1983). However, these systems can be characterized by important conceptual and technical differences. A short-term objective strategy has stronger instructional validity or correspondence between tests and instruction (Yalow & Popham, 1983). The monitoring probes for short-term measurement are related directly to current instructional material; so, for example, if an instructional intervention is introduction of the r-controlled phonics rule, the monitoring measure is reading r-controlled words. Alternately, with a long-term goal approach, the monitoring probes are not related directly to the instructional material. The instructional intervention may be introduction of the r-controlled phonics rule, whereas the monitoring measure may involve oral reading fluency, accuracy, and/or comprehension on second grade passages.

Although a short-term objective approach enjoys stronger instructional validity (Bowers, 1980), a long-term goal strategy possesses at least three advantages. First, it demonstrates better content validity or representation of the ultimate desired performance, i.e., reading fluency/comprehension (see Fuchs & Fuchs, 1985a). Second and relatedly, its concurrent validity or correlation with other measures of achievement is stronger than that of a short-term objective method (Fuchs, 1982). Finally, data analysis is facilitated with a long-term goal approach: Teachers analyze student performance on material representing the same level of difficulty across a long time period, so data analysis can occur across any contiguous portions of a graph. By contrast, trend lines cannot be applied across time within a short-term goal approach when objective mastery occurs

and the meaurement domain simultaneously shifts in difficulty.

In addition to these conceptual and technical differences, these alternative procedures have different practical implications for practitioners. Short-term objective measurement is easier for practitioners to understand and, therefore, teachers seem to prefer it for communicating progress to fellow professionals and parents (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1982). However, it also requires teachers to create new monitoring measures often, as students master the hierarchy of objectives. This frequent change in measurement requires additional time commitments from teachers (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1982).

Therefore, important conceptual, technical, and practical differences are inherent in these two approachers to determining what to measure. Even more importantly, however, critical differences in student achievement outcomes are associated with these alternative monitoring procedures. In a recent meta-analysis, Fuchs and Fuchs (1985a) found that when progress was measured toward long-term goals, effect sizes calculated on global achievement test dependent measures were an average .51 higher than on outcome measures that were similar to the monitoring probes. On the other hand, when progress was measured toward a series of short-term goals, effect sizes were a mean .40 lower on dependent measures that represented global achievement tests than on probe-like measures.

These findings indicate that in order to promote the type of outcome special educators desire (i.e., global growth vs. mastery of discrete curriculum units), goal monitoring methods need to be selected carefully. Specifically, as practitioners develop programmatic or IEP goals and objectives and related curriculum-based monitoring procedures, both the curricular and content validity of their measurement procedures must be addressed. Curricular validity refers to the match between testing and IEP

goals and objectives; content validity, the correspondence between testing and the true domain in which proficiency is desired (Yalow & Popham, 1984). Curricular and content validity are addressed simultaneously only when practitioners write "significant rather than trivial" IEP goals and objectives, which relate well to the true desired outcome performance (Popham et al., 1985). Attention to this dual criterion allows "measurement-driven instruction" (Popham et al., 1985), or ongoing assessment of pupil progress for instructional planning, to assume an important effect on achievement. It implies that practitioners monitor progress toward long-term goals, an approach that appears to promote global effects on student achievement. Practitioners may wish to use this strategy to complement analysis of short-term goal mastery, a system that, on the other hand, can guide instructional programming decisions more directly.

The finding that long-term goal monitoring relates better to global achievement outcome measures may be especially important in the education of remedial and handicapped students, who typically have poorly developed strategies for maintaining and transfering skills (Anderson-Inman, Walker, & Purcell, 1984; White, 1984). Short-term goal measurement focuses on instructionally-related, relatively restricted domains of material for a time period and then, upon mastery of that material, the measurement and instructional focus simultaneously changes. Such a paradigm may be problematic for at least two reasons. First, it may discourage teachers from reviewing material sufficiently to allow for long-term skill maintenance. Second, a close connection between instruction and measurement may encourage teachers to teach new skills to students within the framework of measurement tasks. For example, if the measurement procedure requires the pupil to read consonant-vowel-consonant words from a list, the teachers may focus instruction on reading these words from lists. As noted by

Goodstein (1982), there may be danger in tying the instructional format too closely to the assessment device or of narrowly defining content-x-format domains of criterion-referenced assessment. Such a restricted instructional format may limit generalization of skills. A more global, long-term approach to monitoring may encourage teachers to incorporate instructional procedures that better promote skill maintenance and generalization.

Goal ambitiousness. A third issue relevant to the question of what to measure concerns goal ambitiousness, or the mastery criteria toward which teachers and students strive. Inherent in the nature of special education goals is improvement of student growth rates. Nevertheless, a persistent and ubiquitous problem in goal specification is ambitiousness: Given a current performance level and an academic year's worth of special education instructional opportunity, how ambitiously ought teachers and IEP teams establish student expectations for improvement?

In exploring this question, Fuchs, Fuchs, and Deno (1985) investigated the importance of goal ambitiousness and goal mastery to student achievement. Subjects were 58 mildly to moderately learning disabled, educable mentally retarded, and behavior disordered students, whose special education teachers assessed baseline performance and set reading goals according to a standard format. On the basis of the relation between baseline and the anticipated goal performance, students were assigned to goal ambitiousness groups. For 18 weeks, teachers implemented students' goals. Then, end-of-treatment goal mastery was determined, and pre- and posttest achievement scores were entered into analyses of covariance. These analyses revealed that goal ambitiousness was associated positively with achievement; goal mastery was not.

Although this study was correlational rather than experimental, results provide tentative evidence that when teachers establish moderately

to highly ambitious goals, students achieve better, regardless of whether or not goals actually are attained. Given the movement in recent decades to create "schools without failure" (Glasser, 1969), to provide "riskless" special education (Mann, 1984), and to develop educational goals that insure goal attainment (cf. Clifford, 1984), the finding that goal ambitiousness, not mastery, was associated with achievement may be unsettling. However, an optimal challenge, which increases teachers' and students' persistence, task initiation, and task resumption, may lead to improved task performance (Clifford, 1984), and these constructive effects of striving may be facilitated by factors inherent in many ongoing monitoring methodologies (see Fuchs et al., 1985). These factors include high and concrete goal awareness as well as unambiguous, easily available, and highly detailed assessment and evaluation information (Clifford, 1984). Therefore, it appears that relatively ambitious goals may be an important dimension as teachers determine procedures for monitoring student progress.

Frequency of Measurement

Criterion-referenced formative evaluation involves ongoing collection of student performance data. Yet, the precise frequency with which measurement occurs can vary and must be determined by practitioners. Relevant considerations in making this determination are technical, practical, and effectiveness concerns (Deno, Mirkin, & Fuchs, 1982). Three important respective questions are: What measurement frequency renders reliable, valid, and sensitive representations of student achievement?; What measurement frequency can be employed by a teacher without excessive time demands?; and What measurement frequency relates to improved student growth? Each of these considerations is explored briefly below.

Technical considerations. Criterion-referenced measurement has received increasing attention over the years as an alternative to

traditional, global, normative assessment, because criterion-referenced measures typically are isomorphic with respect to instructional objectives and therefore, have stronger curricular validity. While this strength is appealing to practitioners, it fails to constitute sufficient grounds for psychometric adequacy. In fact, among published criterion-referenced tests, there is scant empirical support for technical strength. Inspection of 12 commercial criterion-referenced tests revealed that only four test manuals addressed reliability or validity at all, and authors of only two instruments investigated more than one aspect of test adequacy (Tindal et al., 1983). Relatedly, empirical analyses of commercial criterion-referenced tests revealed varying indices of reliability and validity, with many estimates falling considerably below acceptable levels (Tindal et al., 1985). Additionally, criterion-referenced assessment frequently requires educators to create their own testing materials and procedures; and given the time-consuming nature of reliability and validity studies, it is infeasible to investigate psychometric characteristics of every teacher-created test.

Thus, the technical acceptability of criterion-referenced measures remains largely unknown. Nevertheless, measurement theory indicates a focus on the methodology of measurement, rather than on the content and format of each test, might result in acceptable reliabilities for criterion-referenced tests. If it were demonstrated that certain measurement methods tend to enhance the acceptability of criterion-referenced measurement, then those methods could be employed with any test to improve reliability.

Frequency of measurement has been identified as an aspect of criterion-referenced assessment methodology that affects reliability. White (1971) established that in order to project a reliable performance trend, a minimum of seven data points was necessary. This finding indicates that to

insure an adequate data base on which to support decisions concerning the efficacy of student programs and to avoid prolonged utilization of inappropriate instructional strategies, practitioners should collect data frequently.

Fuchs, Deno, and Marston (1983) investigated another aspect of technical adequacy related to measurement frequency. Borrowing from the measurement literature demonstrating that behavior averaged over occasions reduces measurement error (Epstein, 1980), they hypothesized that the stability of criterion-referenced measurement should improve as the number of observations over which estimates are aggregated increases. In a series of two experiments, they demonstrated that, for initially imprecise curriculum-based measures of academic proficiency, aggregating estimates of performance over as few as two occasions increased stability well within acceptable levels. Therefore, these findings also support frequent data collection.

Practical concerns. While technical issues seem to support daily measurement, practical considerations suggest a leaner measurement schedule. Evidence (Fuchs, Wesson, Tindal, Mirkin, & Deno, 1981) indicates that preparing for measurement, measuring, scoring performance, recording scores, and putting materials away can be time-consuming: Ten teachers, trained during a series of workshops, initially spent almost 13 minutes per measurement task; after considerable practice in the field, they devoted an average of approximately 2 minutes. These findings were replicated by Rieth and colleagues (see Rieth, 1982). Multiplied across a caseload of 15 students, each of whom is measured on three curriculum tasks, these results suggest that measurement potentially can occupy a significant amount of teacher time. Therefore, a measurement frequency of twice weekly rather than daily may represent a reasonable compromise: Teacher measurement time

is reduced while technical properties may be maintained.

Effectiveness concerns. In further support of a twice weekly measurement schedule, research indicates that no additional benefits accrue to student achievement as a function of increasing measurement frequency beyond twice weekly. In a quantitative synthesis of relevant controlled studies, L. Fuchs and D. Fuchs (in press) found that the average effect sizes associated with measurement that occurs twice weekly, three times weekly, and daily, respectively, were .85, .41, and .69, with no related statistically significant difference.

Therefore, the available research base indicates that daily measurement may generate the most technically adequate data base. However, the literatures addressing practical and effectiveness considerations, respectively, suggest that (a) practitioners' time constraints mitigate against daily measurement, and (b) student achievement effects associated with daily and twice weekly measurement are comparable. These findings suggest that practitioners may wish to collect student performance data twice weekly.

Data Display

For measurement systems in which time-series data are essential, such as ongoing monitoring of student progress and applied behavior analysis, agreement prevails that graphing is critical: It assists in organizing data for formative evaluation, provides a detailed numerical summary and visual description of performance, and facilitates communication of program results (Tawney & Gast, 1984). Moreover, available research on the effectiveness of ongoing monitoring systems indicates achievement is associated with graphed displays. When data are charted rather than simply recorded, achievement improves approximately .50 of a standard deviation unit (L. Fuchs & D. Fuchs, in press).

Despite concurrence on the importance of graphing and empirical data to support its effectiveness, important differences exist concerning specific graphing conventions, the most salient of which may be the type of graphing paper employed. Some programs advocate the use of ratio or logarithmically scaled graph paper (e.g., Lindsley, 1977; White & Haring, 1980), where the rate scale is adjusted to display proportional changes in student behavior. For example, the distance from 10 to 20 is identical to the change from 20 to 40 or from 40 to 80. In contrast, developers of other monitoring systems support the use of equal interval, or conventional, graphing paper (e.g., Deno & Mirkin, 1977). Below, the related controversy is reviewed.

Logarithmically scaled paper has been described as technically superior by its proponents (see, for example, White & Haring, 1980), because the ratio scale is supposed to reflect the proportional way in which natural change occurs more accurately than equal interval paper. Additionally, logarithmic graphing paper has been described as more feasible than equal interval paper because, given the large behavior range incorporated in one graph, a single chart can be used to display all relevant behaviors and can be used to make comparisons among different behaviors (White & Haring, 1980).

On the other hand, equal interval graphing may facilitate data analysis (Tawney & Gast, 1984), and has been characterized frequently as easier for students and teachers to understand. Some propose that this understanding and relative ease in data analysis may, in turn, result in more consistent implementation of monitoring systems (Mirkin, Fuchs, & Deno, 1982). Additionally, Marston (1982) explored the prediction capabilities of the two graphing methods and found that trend lines on equal interval paper predicted future performance more accurately than did trend lines drawn on

ratio scaled paper. This controversy surrounding relative merits associated with the graphing methods continues. Nevertheless, except for the Marston study, no empirical contrast of technical or teacher and student concerns has been identified. Therefore, the objective data base is inadequate to support the technical or logistical strengths of either graphing approach.

Despite continuing discussion of which type of graphing method is technically and logistically superior, relatively little attention has been directed toward investigation of which method is associated with better student achievement. In one relevant identified report, Branstetter and Merz (1978) conducted a series of two experiments, in which they compared gains made while charting scores on linear graphs with those made while simply recording raw scores, and then compared gains associated with charting scores on ratio scaled graphs with those related to simple recording. Unfortunately, no attempt was made to contrast the effectiveness of graphing on linear and ratio scaled graphs. Furthermore, the children employed in the two studies were neither randomly assigned nor similar to each other, rendering it impossible to draw valid comparisons between the investigations.

In an attempt to explore the question of how graphing method contributes to student achievement, Fuchs and Fuchs (1985b) conducted a meta-analysis of available controlled studies on the effectiveness of ongoing monitoring systems, coding studies by graphing convention and then computing and comparing effect sizes for studies in which equal interval and ratio scaled papers were employed. Results indicated that graphing methods did not produce a statistically reliable difference in student achievement. Moreover, the difference between the mean effect sizes of .2 standard deviation unit represented a difference of little practical importance (Cohen, 1977). Therefore, student achievement effects do not appear to

provide a basis on which to select a graphing convention.

In sum, currently available data support the use of graphing, but fail to provide a basis for selecting a specific graphing convention. Little available information concerns technical properties; the data base on relative logistical strengths and weaknesses of the methods is scant; and the literature on effects on student achievement reveals no reliable effects. Consequently, as practitioners design criterion-referenced formative evaluation procedures for monitoring their pupils' attainment of goals, they might consider graphing as an essential procedural element, but rely on individual preferences and logistical considerations for deciding between equal interval and ratio scaled paper for graphed displays.

Data Utilization

Although teachers may collect student performance data according to designated time schedules, they frequently fail to employ those data meaningfully to develop students' educational programs (Baldwin, 1976; White, 1974). For example, Tindal, Fuchs, Mirkin, Christenson, and Deno (1981) found that teachers often maintained instructional programs long periods despite that student performance data clearly indicated those programs were not producing student improvement. To complicate data utilization further, although teachers may analyze data correctly to recognize when interventions are not effective, they experience difficulty generating substantively important modifications in their students' programs (Fuchs, Deno, & Mirkin, 1982).

Examples of data-utilization procedures are provided in the work of Haring and White and their colleagues as well as that of Deno and associates. Haring, White, and Liberty (1979) developed a set of rules entitled "Experimental Data-Decision Rules with Minimal 'Celeration." These rules require practitioners to assess student performance in relation to an

aimline, which connects the baseline date and level of performance with the goal date and anticipated performance level. In her study of the effectiness of such rules, Martin (1980) adapted Haring et al.'s data-utilization strategy as follows: (a) If a student's performance was on the aimline on one day and within five words of the aimline on the next day, the teacher progressed the student to the next curriculum step; (b) If a student's correct performance was above the aimline for five consecutive days, the teacher drew a new aimline parallel to but above the original one; and (c) If a student's correct performance was below the aimline on two consecutive days, the teacher introduced a change in the instructional format, wherein data trends were used to determine the type of instructional problem (i.e., inappropriate instructional step or problems of compliance, fluency-building, acquisition, or format). The type of instructional problems then dictated general strategies concerning the nature of the instructional change.

In her experiment, Martin compared effects on student achievement among groups that (a) collected but did not graph daily data, (b) graphed and employed the above rules to determine what and when to change instructional programs, and (c) graphed data and employed the "when to change" rules without using the "what to change" rules. Results indicated that posttest scores of the second and third groups were significantly higher than those of the first group on certain measures, with no significant difference between the two data-utilization rule groups. Unfortunately, given that graphing, in and of itself, positively affects student achievement (L. Fuchs & D. Fuchs, in press), this study is difficult to interpret: It provides inadequate control for separating the effects of data-utilization from those of simply graphing. However, results do suggest that the use of general rules for specifying the nature of changes may not

result in additional achievement gains over the extent of improvement associated with using "when to change" rules.

Deno and his colleagues have employed two contrasting data-utilization procedures. One strategy, a "therapeutic approach," is aimline-referenced, the other of which, an "experimental" approach, is referenced to the level of performance in the immediately preceding instructional phase. With the therapeutic approach, data interpretation consists of the application of the following rule: If on 7 to 10 consecutive data points, the student performance trend is below the aimline, then an instructional change is introduced. With the experimental approach, the trend, level, variability, and step change in students' performance data are analyzed every 7 to 10 consecutive data points, and compared to the same indices calculated in the preceding phase. When data analysis occurs in an experimental approach, program change is introduced: Data analysis does not determine <u>whether</u> change occurs, but rather facilitates determination of <u>what</u> to change. If the current program element is relatively ineffective, then it is dropped and a new programmatic element is initiated; if relatively effective, then it is maintained, but a new program element nonetheless is introduced in an attempt to boost the performance level even further.

Although studies comparing these data-utilization systems to each other as well as to those of Haring and his associates are scant and inconclusive, investigations document the effectiveness of each rule-based approach separately (Mirkin, Deno, Tindal, & Kuehnle, 1980; Deno, 1985). Similarly, quantitative synthesis of controlled studies of various types of data-utilization systems fail to provide evidence for the differential effectiveness of one type of data-utilization method. Nevertheless, integration of findings indicates persuasive support for the effectiveness

of data-utilization rules in general. L. Fuchs and D. Fuchs (in press) found that, on average, systematic formative evaluation that incorporated evaluation rules increased student achievement approximately .5 of a standard deviation unit over systematic formative evaluation without such rules. With rules, the mean effect size was .91, indicating that the upper 50% of the experimental group distribution, wherein evaluation rules were employed, exceeded approximately 82% of the control group (no systematic formative evaluation) distribution.

Conclusions and Delineation of Research Questions

Available research supports the use of ongoing criterion-referenced monitoring systems to improve the instructional programs of mildly handicapped students. Research also provides the basis for several statements on the nature of effective monitoring systems among mildly handicapped populations. Specifically, the data base supports the use of (a) certain empirically validated critical behaviors as the focus of monitoring activity, (b) long-term goal statements that may encourage teachers to focus not only on the immediate instructional content, but also on maintenance and generalization of skills, and (c) relatively ambitious goals that support task persistence and striving. The literature also provides rationale for teachers to measure student performance at least twice weekly, to graph data using their preferred convention, and to employ data-utilization rules for determining when, and perhaps how, to modify students' instructional programs.

Nevertheless, the same literature leaves many critical questions unanswered and constitutes the basis for delineating a research program in the area of monitoring the progress of mildly handicapped pupils. Among the

issues requiring additional clarification and empirical exploration, several concern teachers' apparent difficulty in implementing monitoring systems and in using data meaningfully to develop instructional programs (Baldwin, 1976; Rieth, 1982; Tindal et al., 1981; White, 1974). Related research questions include: (a) What antecedent and consequential conditions (such as type of training, system level support, and professional feedback) increase the probability that teachers will measure and evaluate student academic performance according to designated schedules?; and (b) Does computer technology, designed to complement monitoring activities by facilitating data collection, storage, graphing, analysis, and evaluation (see Hasselbring, 1985; Hasselbring & Hamlett, 1983), affect the rate and accuracy with which monitoring procedures are employed and does it improve teachers' instructional behavior and/or pupil achievement?

Additional questions concern dimensions of useful monitoring systems. Related research questions include the following:

(a) What is the effect of goal ambitiousness levels, contrasted within a well-controlled experiment, on teacher decisionmaking and on student achievement?

(b) What are the technical and practical effects and student achievement outcomes associated with the use of six-cycle and equal interval paper, when these graphing conventions are contrasted within an experimental study?

(c) How does the use of data-utilization rules affect teacher decisionmaking?

(d) What are the differential effects of graphing data, using "when to change" data-utilization rules, and employing "how to change" data-utilization rules on student achievement and teachers' instructional behavior?

(e) How do different data-utilization rules, such as experimental and therapeutic decisionmaking, relate to student achievement?

(f) How can expert systems be incorporated with computer technology to facilitate teachers' determination of how to modify students' instructional porgrams, and what are the effects of using such expert systems on student achievement and teachers' pedagogical behavior?

(g) What are the effects of computer printout graphed displays on teacher decisionmaking and student achievement?

These represent a handful of many useful questions that provide interesting territory for well conceptualized and designed investigation. In light of evidence clearly indicating the efficacy of employing systematic ongoing criterion-referenced monitoring systems, such continued development and research appears to be potentially important to educators of mildly handicapped and remedial students.

References

Anderson-Inman, L., Walker, H., & Purcell, J. (1984). Promoting the transfer of skills across settings: Transenvironmental programming for handicapped students in the mainstream. In W.L. Heward, T.E. Heron, D.S. Hill, & J. Trap-Porter (Eds.), Focus on behavior analysis in education (pp. 17-37). Columbus, OH: Merrill.

Arter, J.J., & Jenkins, J.R. (1979). Differential diagnosis-prescriptive teaching: A critical appraisal. Review of Educational Research, 49, 517-555.

Baldwin, V. (1976). Curriculum concerns. In M.A. Thomas (Ed.), Hey, don't forget about me. Reston, VA: Council for Exceptional Children.

Bloom, B.S., Hastings, J.T., & Madaus, G.F. (1971). Handbook of formative and summative evaluation of student learning. New York: McGraw Hill.

Bobbitt, F. (1918). The curriculum. Boston: Houghton Mifflin.

Bowers, J.E. (1980). End-of-year evaluation report for the 1979-1980 Title I Neglected and Delinquent Program of the Minneapolis Public Schools. Unpublished manuscript. (Available from Educational Services Division, Minneapolis Public Schools, 807 N.E. Broadway, Minneapolis, MN 55413.)

Brandstetter, G., & Merz, C. (1978). Charting scores in precision teaching for skill acquisition. Exceptional Children, 45, 42-48.

Campbell, R.F., Cunningham, L.L., Nystrand, R.O., & Usdan, M.D. (1975). The origin and control of American schools (3rd ed.). Columbus, OH: Merrill.

Clifford, M.M. (1984). Thoughts on a theory of constructive failure. Educational Psychologist, 19, 108-120.

Cohen, J. (1977). Statistical power analysis for the behavioral sciences. New York: Academic Press.

Coles, G.S. (1978). The learning-disabilities test battery: Empirical and social issues. Harvard Educational Review, 3, 313-340.

Deno, S.L. (1985). Personal communication.

Deno, S.L., Marston, D., & Mirkin, P.K. (1982). Valid measurement procedures for continuous evaluation of written expression. Exceptional Children, 48, 368-371.

Deno, S.L., & Mirkin, P.K. (1977). Data-based program modification: A manual. Reston, VA: Council for Exceptional Children.

Deno, S.L., Mirkin, P.K., & Chiang, B. (1982). Identifying valid measures of reading. Exceptional Children, 49, 36-45.

Deno, S.L., Mirkin, P.K., & Fuchs, L.S. (1982). Decision framework for developing measurement and evaluation systems. In P.K. Mirkin, L.S. Fuchs, & S.L. Deno (Eds.), Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20, pp.17-29). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Deno, S.L., Mirkin, P.K., Lowry, L., & Kuehnle, K. (1980). Relationships among simple measures of written expression and performance on standardized achievement tests (Research Report No. 22). Minneapolis: University of Minnesota Institute for Learning Disabilities. (ERIC Document Reproduction Service No. ED 197 508)

Dunn, L. (1968). Special education for the mildly retarded -- Is much of it justified? Exceptional Children, 35, 5-24.

Eisner, E.W. (1967). Educational objectives: Help or hindrance. School Review, 75, 251-282.

Epstein, S. (1980). The stability of behavior, II: Implications for psychological research. American Psychologist, 35, 790-806.

Eubanks, E.E., & Levine, D.U. (1983). A first look at effective schools projects in New York City and Milwaukee. Phi Delta Kappan, 64, 697-702.

Fuchs, D., & Fuchs, L.S. (in press a). Linking assessment to instructional intervention: An overview. In L.S. Fuchs & D. Fuchs (Eds.), Linking assessment

to instructional intervention: A special issues. School Psychology Review.

Fuchs, D., & Fuchs, L.S. (in press b). Test procedure bias: A meta-analysis of examiner familiarity effects. Review of Educational Research.

Fuchs, D., Fuchs, L.S., Power, M.A., & Dailey, A. (1985). Bias in the assessment of handicapped children. American Educational Research Journal, 22, 185-198.

Fuchs, L.S. (1981). The concurrent validity of progress measures of basal reading material. Unpublished doctoral dissertation, University of Minnesota.

Fuchs, L.S. (1982). Reading. In P.K. Mirkin, L.S. Fuchs, & S.L. Deno (Eds.), Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20, pp. 29-74). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities.

Fuchs, L.S., Deno, S.L., & Marston, D. (1983). Improving the reliability of curriculum-based measures of academic skills for psychoeducational decision making. Diagnostique, 8, 135-149.

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1982). The effect of continuous evaluation of student goals on teacher decisions (Research Report No. 97). Minneapolis: University of Minnesota Institute for Research on Learning Disabilties. (ERIC Document Reproduction Service No. ED 226 043)

Fuchs, L.S., Deno, S.L., & Mirkin, P.K. (1984). The effects of frequent curriculum-based measurement and evaluation on pedgagogy, student achievement, and student awareness of learning. American Educational Research Journal, 21, 449-460.

Fuchs, L.S., & Fuchs, D. (1984). Criterion-referenced assessment without measurement: How accurate for special education? Remedial and Special Education, 5, 29-32.

Fuchs, L.S., & Fuchs, D. (1985a). Curriculum-based assessment of progress toward long- and short-term goals. Unpublished manuscript. (Available from L.S. Fuchs, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203.)

Fuchs, L.S., & Fuchs, D. (1985b). The relation between methods of graphing student performance data and student achievement. Unpublished manuscript. (Available from L.S. Fuchs, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203.)

Fuchs, L.S., & Fuchs, D. (in press). Effects of systematic formative evaluation: A meta-analysis. Exceptional Children.

Fuchs, L.S., Fuchs, D., & Deno, S.L. (1985). The importance of goal ambitiousness and goal mastery to student achievement. Exceptional Children, 52, 63-71.

Fuchs, L.S., Fuchs, D., & Warren, L.M. (1982). Special education practice in evaluating student progress toward goals (Research Report No. 82). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 224 198)

Fuchs, L.S., Wesson, C., Tindal, G., Mirkin, P.K., & Deno, S.L. (1981). Teacher efficiency in continuous evaluation of IEP goals (Research Report No. 53). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 215 467)

Fuchs, L.S., Wesson, C., Tindal, G., Mirkin, P.K., & Deno, S.L. (1982). Instructional changes, student performance, and teacher preferences: The effects of specific measurement and evaluation procedures (Research Report No. 64). Minneapolis: Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. 218 849)

Gersten, R., Carnine, D., & White, W.A.T. (1984). The pursuit of clarity: Direct instruction and applied behavior analysis. In W.L. Heward, T.E. Heron, D.S. Hill, & J. Trap-Porter (Eds.), Focus on behavior analysis in education (pp. 38-57). Columbus, OH: Merrill.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. American Psychologist, 36, 923-936.

Glass, G. V. (1983). Effectiveness of special education. Policy Studies Review, 2

(Special No. 1), 65-78.

Glasser, W. (1969). Schools without failure. New York: Harper & Row.

Goodman, L. (1985). The effective schools movement and special education. Teaching Exceptional Children, 17, 102-105.

Goodstein, H.A. (1982). The reliability of criterion-referenced tests and special education: Assumed versus demonstrated. Journal of Special Education, 16, 37-48.

Hammill, D.D., & Larsen, S. (1974). The effectiveness of psycholinguistic training. Exceptional Children, 41, 5-14.

Hammill, D.D., & Wiederholt, J.L. (1973). Review of the Frostig visual perception test and the related training program. In L. Mann & D.A. Sabatino (Eds.), First review of special education (Vol. 1, pp. 33-48). Philadephia: JSE Press.

Haring, N.G., White, O.R., & Liberty, K.A. (1979). Field initiated research studies: An investigation of learning and instructional hierarchies in severely and profoundly handicapped children: Annual Report 1978-1979. Seattle: University of Washington, Child Development and Mental Retardation Center, Experimental Education Unit.

Hasselbring, T.S (1985). Toward the development of expert assessment systems. Working paper. (Available from T. Hasselbring, Box 328, Peabody College, Vanderbilt University, Nashville, TN 37203.)

Hasselbring, T.S., & Hamlett, C.L. (1983). AIMSTAR: A data management and decision-making system. Portland: ASIEP Education Co.

Hoffman, J.V., & Rutherford, W.L. (1984). Efffective reading programs: A critical review of outlier studies. Reading Research Quarterly, 20, 79-92.

Lindsley, O. (1964). Direct measurement and prosthesis of retarded behavior. Journal of Education, 147, 62-81.

Lindsley, O. (1971). Precision teaching in perspective: An interview with Ogden R. Lindsley. Teaching Exceptional Children, 3 (3), 114-119.

Lovitt, T.C. (1977). Inspite of my resistance, I've learned from children. Columbus, OH: Merrill.

Lovitt, T.C. (1981). Notes on behavior modification. Journal of Special Education, 15, 395-400.

Mann, L. (1984). On seeking a riskless education. Journal of Special Education, 18, 94-97.

Marston, D. (1982). The technical adequacy of direct, repeated measurement of academic skills in low achieving elementary students. Unpublished doctoral disseration, University of Minnesota.

Marston, D., & Deno, S.L. (1981). The reliability of simple, direct measures of written expression (Research Report No. 50). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 212 663)

Marston, D., Fuchs, L.S., & Deno, S.L. (1985). Measuring pupil progress: A comparison of standardized achievement tests and curriculum-based measures. Unpublished paper. (Available from D. Marston, 2752 41st Ave. S.E., Minneapolis, MN 55406.)

Martin, M.A. (1980). A comparison of variations in data utilization procedures on the reading performance of mildly handicapped students. Unpublished doctoral disseration, University of Washington.

Mirkin, P.K., Deno, S.L., Tindal, G., & Kuehnle, K. (1980). Formative evaluation: Continued development of data utilization systems (Research Report No. 23). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 197 510)

Mirkin, P.K., Fuchs, L.S., & Deno, S.L. (Eds.). (1982). Considerations for designing a continuous evaluation system: An integrative review (Monograph No. 20). Minneapolis: University of Minnesota Institute for Research on Learning Disabilites.

Mulhern, J. (1959). A history of education (2nd ed.). New York: Ronald Press.

Peterson, D.L., Albert, S.C., Foxworth, A.M., Cox, L.S., & Tilley, B.K. (1985). Effective schools for all students: Current efforts and directions. Teaching Exceptional Children, 17, 106-111.

Popham, W.J. (1980). Domain specification strategies. In R.A. Berk (Ed.), Criterion-referenced measurement: The state of the art. Baltimore: The Johns Hopkins University Press.

Popham, W.J., Cruse, K.L., Rankin, S.C., Sandifer, P.D., & Williams, P.L. (1985). Measurement-driven instruction: It's on the road. Phi Delta Kappan, 66, 628-634.

Rieth, H.J. (1982). Evaluating and providing feedback on the effectiveness of instruction for handicapped children integrated in inner-city secondary schools: Final Report, Grant No. NIE-G-80-0089. Bloomington, IN: Center for Innovation in Teaching the Handicapped, Indiana University.

Rieth, H.J., Polsgrove, L., & Semmel, M.I. (1981). Instructional variables that make a difference: Attention to task and beyond. Exceptional Education Quarterly, 2 (3), 61-72.

Salvia, J., & Ysseldyke, J. (1985). Assessment in special and remedial education (3rd ed.). Boston: Houghton-Mifflin.

Tawney, J.W., & Gast, D.L. (1984). Single subject research in special education. Columbus, OH: Merrill.

Tindal, G., Shinn, M., Fuchs, L.S., Fuchs, D., Deno, S.L., & Germann, G. (1983). The technical adequacy of a basal reading series mastery test (Research Report No. 113). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 236 191)

Tindal, G., Fuchs, L.S., Fuchs, D., Shinn, M.R., Deno, S.L., & Germann, G. (1985). Empirical validation of criterion-referenced tests. Journal of Educational Research, 78, 203-209.

Tindal, G., Fuchs, L.S., Christenson, S., Mirkin, P.K., & Deno, S.L. (1981). The relationship between student achievement and teacher assessment of short- or long-term goals (Research Report No. 61). Minneapolis: University of Minnesota Institute for Research on Learning Disabilities. (ERIC Document Reproduction Service No. ED 218 846)

White, O.R. (1971). A pragmatic approach to the description of progress in the single case. Unpublished doctoral disseration, University of Oregon.

White, O.R. (1974). Evaluating the educational process. Seattle: University of Washington, Child Development and Mental Retardation Center, Experimental Education Unit, 1974.

White, O.R. (1984). Descriptive analysis of extant research literature concerning skill generalization and the severely/profoundly handicapped. In M. Boer (Ed.), Investigating the problem of skill generalization: Literature review (pp. 1-19). Seattle: University of Washington, Washington Research Organization.

White, O.R., & Haring, N.G. (1977). Exceptional teaching. Columbus, OH: Merrill.

White, O.R., & Haring, N.G. (1980). Exceptional teaching (2nd ed.). Columbus, OH: Merrill.

Yalow, E.S., & Popham, W.J. (1983). Content validity at the crossroads. Educational Researcher, 12 (8), 10-14, 21.

Ysseldyke, J.E. (1979). Psychoeducational assessment and decision making. In J.E. Ysseldyke & P.K. Mirkin (Eds.), Proceedings of the Minnesota Roundtable Conference on Assessment of Learning Disabled Children (Monograph No. 8, pp. 15-36). Minneapolis: University of Minnesota, Institute for Research on Learning Disablities.

Ysseldyke, J.E., & Thurlow, M.L. (1984). Assessment practices in special education: Adequacy and appropriateness. Educational Psychologist, 19, 123-136.